Journal of Biomedical Informatics 54 (2015) 213–219

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx

Saeed Mehrabi [a,d], Anand Krishnan [a], Sunghwan Sohn [d], Alexandra M. Roch [b], Heidi Schmidt [b], Joe Kesterson [c], Chris Beesley [c], Paul Dexter [c], C. Max Schmidt [b], Hongfang Liu [d,*], Mathew Palakal [a,*]

[a] School of Informatics and Computing, Indiana University, Indianapolis, IN, USA
[b] Department of Surgery, Indiana University, Indianapolis, IN, USA
[c] Regenstrief Institute, Indianapolis, IN, USA
[d] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

ARTICLE INFO

Article history:
Received 16 September 2014
Revised 22 January 2015
Accepted 24 February 2015
Available online 16 March 2015

Keywords:
Natural language processing
Dependency parser
Negation

ABSTRACT

In Electronic Health Records (EHRs), much of valuable information regarding patients' conditions is embedded in free text format. Natural language processing (NLP) techniques have been developed to extract clinical information from free text. One challenge faced in clinical NLP is that the meaning of clinical entities is heavily affected by modifiers such as negation. A negation detection algorithm, NegEx, applies a simplistic approach that has been shown to be powerful in clinical NLP. However, due to the failure to consider the contextual relationship between words within a sentence, NegEx fails to correctly capture the negation status of concepts in complex sentences. Incorrect negation assignment could cause inaccurate diagnosis of patients' condition or contaminated study cohorts. We developed a negation algorithm called DEEPEN to decrease NegEx's false positives by taking into account the dependency relationship between negation words and concepts within a sentence using Stanford dependency parser. The system was developed and tested using EHR data from Indiana University (IU) and it was further evaluated on Mayo Clinic dataset to assess its generalizability. The evaluation results demonstrate DEEPEN, which incorporates dependency parsing into NegEx, can reduce the number of incorrect negation assignment for patients with positive findings, and therefore improve the identification of patients with the target clinical findings in EHRs.

1. Introduction

Electronic health records (EHRs) contain valuable clinical information that can be used for various applications such as clinical decision support systems, medication reconciliation, public health emergency surveillance, and quality measurements [1]. However these applications are not readily feasible because much of the information in EHR is in free text format. Natural language processing (NLP) systems have been developed to extract clinical concepts from text, yet this is not an easy task because the meaning of a concept is significantly affected by modifiers such as negation. Negative clause is defined as "an assertion that some event, situation, or state of affairs does not hold. Negative clauses usually occur in the context of some presupposition, functioning to negate or counter-assert that presupposition" [2].

A study of negation has shown that clinical observations are frequently negated in clinical narratives [3]. Negation detection in clinical language tends to be very trivial in sentences such as "no fracture", "patient denies headache", and "she does not have marked dysmenorrhea.". Therefore simplistic approaches such as NegEx [4] that use negation cue words without considering the semantic of a sentence perform well. However, the simplistic approaches sometimes fail to correctly identify the negation status of clinical concepts in sentences with complex structure. We have faced with this problem while using NegEx in our NLP system that automates the identification and tracking of patients with pancreatic cysts [5]. Table 1 shows some examples of such sentences where NegEx incorrectly negates pancreatic cyst concepts.

Aiming to reduce the number of missing pancreatic cyst patients in our NLP system inspired us to improve the negation assignment of NegEx by incorporating dependency parsing into NegEx. Dependency relation is a binary asymmetric relation

* Corresponding authors at: Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: +1 507 773 0057 (H. Liu). 535W. Michigan St, Indianapolis, IN 46202, USA. Tel.: +1 317 278 7689; fax: +1 317 278 7669 (M. Palakal).
E-mail addresses: liu.hongfang@mayo.edu (H. Liu), mpalakal@iupui.edu (M. Palakal).

http://dx.doi.org/10.1016/j.jbi.2015.02.010
1532-0464/

Table 1
Examples of sentences where NegEx failed to capture the correct negation status of concepts denoted by bold letters.

Record type	Sample sentence
Discharge summary	Additionally, there was no evidence of extension of his infected **pseudocyst** into the psoas muscle
Abdomen CT	There is no significant interval change in the 2 large pancreatic **pseudocysts**
OPERATIVE REPORT	We confirmed no evidence of epithelium consistent with a **pseudocyst**
Consultation	Acute pancreatitis with pseudocyst, with no obvious complications of the **pseudocyst** at this point in time
Liver CT W Contr	Although there is no discretely visualized or abnormal enhancing pancreatic mass, there is marked **pancreatic duct dilatation** with side duct ectasia and abrupt cutoff of the pancreatic duct within the pancreatic head

between tokens within a sentence that has been shown to improve various NLP tasks including information extraction [6], negation detection [7], entity disambiguation [8] and many others [9].

We developed and tested our negation identification algorithm focusing on only pancreatic cyst concepts using a single institution data set. In order to evaluate its performance on other clinical concepts and dataset, we applied our system on 159 clinical notes from Mayo Clinic where clinical findings such as disorders and signs/symptoms have been annotated. We compared the performance of our algorithm on Mayo Clinic dataset with NegEx.

2. Related work

Negation detection has been the main or sub task of several challenges in NLP. Assertion classification was one of the three tasks in the 2010 i2b2/VA shared task where each medical concept had to be classified into one of six categories of "*present*", "*absent*", "*possible*", "*conditional*", "*hypothetical*", and "*not associated with the patient*" [10]. Processing modality and negation was the main task of Question Answering or Machine Reading Evaluation (QA4MRE) lab at CLEF 2011 [11]. Negation and speculation in NLP (NeSp-NLP 2010) [12], identifying hedges and their scope in CoNLL-2010 shared task [13], and SEM 2012 shared task of resolving the scope and focus of negation [14] are few other initiatives that show the growing importance of negation processing in the NLP research community.

Corpora used in 2010 i2b2/VA and CoNLL-2010 shared tasks are available to researcher with signing a data use agreement to facilitate the development and evaluation of clinical NLP algorithms. BioScope corpus that was used as part of the CoNLL-2010 shared task has been created by annotating negation and uncertainty in biomedical texts is also publicly available [15]. BioScope corpus consists of clinical text, abstract and full text of scientific articles. The free text clinical notes of BioScope corpus are the radiology reports from the 2007 ICD9 challenge of the Cincinnati children hospital [16]. NegEx has released a de-identified physician annotated test set of 2376 sentences from 120 clinical reports. Also an instruction on how to produce an annotation guideline for biomedical corpus with negation layer is available [17]. Below we review some of the work presented in these challenges or developed outside of theses shared tasks.

In negation detection, rule based techniques have been shown to be effective and widely used in many NLP systems [18,19]. Rule based negation systems can be token-based (e.g., NegEx [4], NegExpander [20], NegFinder [21], NegHunter [22]) ontology-based [23], or utilize syntactic parsing results (e.g., DepNeg [24], ChartIndex [25], Ballesteros et al. [26]). For example, NegEx processes one sentence at a time by finding negation and termination terms. Termination terms are conjunctions such as "*but*" that end the scope of negation terms. There are three types of negation in NegEx algorithm, pseudo negation terms that are similar to negation terms but do not negate clinical conditions, pre-condition negation terms that appear before the clinical findings, and post-condition negation terms that appear after the clinical findings. If a pseudo negation term is found, NegEx skips to the next negation term in the sentence and uses corresponding regular expressions based on pre/post negation terms. NegEx has been extended into an algorithm called ConText in order to determine if a clinical condition of interest is hypothetical, historical or experienced by someone other than patient in addition to negation identification [27]. Both NegEx and ConText have been translated into other languages [28,29].

There are some attempts to incorporate syntactic parsing to improve the negation detection [24,26]. For example, DepNeg is a dependency parser-based negation algorithm that utilizes the dependency structure of a target named entity in the sentence instead of a fixed negation scope [24]. DepNeg uses manual negation rules based on the patterns of dependency paths between the focus (i.e., named entity) and the potential negation terms in the text that enables correctly identifying problematic negations in the traditional negation algorithm, such as NegEx. Similarly, Ballesteros et al. used Minipar dependency parser to determine the scope of negation terms by traversing the dependency path from sentence's verb toward the end of the sentence. They could detect negation terms and their scope in clinical text of BioScope corpus with precision and recall of 0.958 and 0.906 respectively [26].

Machine learning has also been applied in negation detection. For instance, there are twenty-one systems developed for i2b2/VA assertion classification task where majority of them applied various machine learning algorithms including support vector machines (SVMs). The best system achieved 0.9326 micro-averaged *F*-measure using a 2-step approach. Where, in the first step, each word was represented as a feature vector consisting of n-gram, token category, and window of four tokens before and after the word, etc. and then a set of different classifiers were used to predict a score per class for each concept. In the second stage a multi-class SVM was used to predict the final assertion prediction for each token [30]. Similar 2-step approach was applied to BioScope corpus by Diaz et al. where each token in a sentence was classified as negation/speculation signal and a second classifier was used at a sentence level to determine the negation status of concept [31]. Goldin and Champan compared Naïve Bayes and decision trees with default NegEx rule on 207 sentences of clinical records with negation "not". The default NegEx rule negates any UMLS concept within six-word window of "not." Naïve Bayes performed better than decision tree and NegEx [32].

Features used in machine learning algorithms may include results from rule-based systems as well as syntactic parsing results. For example, Grouin et al. used SVM with NegEx and ConText dictionaries before or after a concept in a 5-word window [33]. Wu et al. [34] also used SVM with following list of features, 1) binary feature indicating if a given word appeared in a window size of 3, 5 or 10 from the named entity 2) token in an exact distance from the named entity 3) negation terms 4) DepNeg dependency rules indicating whether a named entity is on the same dependency path as the negation word 5) constituency tree fragments to represent if a named entity is inside a phrase. They trained and test their system on four different corpora of SHARP NLP [35], 2010 i2b2/VA, MiPACQ [36], and NegEx test sets and compared their system with YTEX [37] implementation of NegEx algorithm. Their results were mixed and non conclusive, NegEx performed very well on NegEx test set (*F*-measure = 0.953) but the performance declined on other corpora with lowest *F*-measure of 0.623. Using a single versus all corpora for training the SVM has

also generated mixed results that can be contributed to the diversity of their corpora.

As majority of the systems reviewed above are not publicly available, it is not feasible to compare various systems reported in the literature. Determining the scope of negation is a main challenge in most of rule based methods such as NegFinder that use a context free grammar parser especially when the distance between negation term and concept is more than a few words. For instance in the sentence "*Based on this, he required no operative intervention for his pseudocyst.*" Because of the negation term "no" NegEx will consider the concept "pseudocyst" as negated while "no" is associated with "operative intervention" and not the "pseudocyst". DepNeg attempts to remove this deficiency using dependency parser and shows promising preliminary results while using a limited set of rules on 159 Mayo clinical notes. DepNeg was compared with cTAKES adoption of NegEx, which is customized to Mayo Clinic data. cTAKES is an open source natural language processing tool for information extraction from medical records developed by Mayo Clinic and released under Apache license [18]. DepNeg focused on improving the precision of NegEx therefore it decreased the number of false positives in comparison to cTAKES negation (cTAKES negation-FP: 34, DepNeg-FP: 6) but increased the number of false negatives (cTAKES negation-FN: 47, DepNeg-FN: 61) [24].

There are two approaches of graph-based and transition-based in dependency parser. DepNeg uses ClearParser [38], which is a graph-based dependency parser to determine whether the negation words are on the same path as clinical concepts and therefore negated. Unlike DepNeg, we use a transition-based dependency parser to find if there is any dependency relation between negation words and concepts. And because NegEx had low number of false negatives (high recall) in our training set, we only applied the dependency parser to concepts that are considered negated by NegEx unlike DepNeg that applies dependency parser to all sentences containing negation tokens.

3. Material and methods

This study was conducted under approved institutional review board at each institution.

3.1. Patient cohorts

3.1.1. Indiana university dataset

Longitudinal health records including discharge summary, surgical pathology document, imaging reports (abdominal MRI, CT with/without contrast, Ultrasound, etc.) and other clinical notes (procedure notes, visit notes, letter, consultation, etc.) of patients who visited the Sidney & Lois Eskenazi Hospital in Indianapolis, Indiana was used in this study. The Eskenazi Hospital is a 316-bed hospital providing a comprehensive range of primary and specialty care services in central Indiana. It is comprised of providers who are faculty and residents of the Indiana University (IU) school of medicine. The data was divided into two sets of training data of 664 patients consisting of 1136 reports with 1728 sentences with pancreatic cyst concept and test set of 452 patients with 793 reports and 1462 sentences.

3.1.2. Mayo Clinic dataset

A set of 159 clinical notes with manual annotation of named entities and their negation status by four domain experts was used [39]. There are total of 1007 disorders with 426 unique UMLS concepts and 439 signs and symptoms with 129 unique UMLS concepts.

3.2. DEpEndency ParsEr Negation (DEEPEN)

DEEPEN evaluates concepts that are considered negated by NegEx algorithm; so if a concept is considered affirmed by NegEx, no action is taken. Stanford Dependency Parser (SDP) [40] is applied to sentences containing the negated concept. SDP comprises of 53 grammatical relations (e.g. det: determiner, infmod: infinitival modifier, etc.) that will be generated for words within a sentence [41]. The SDP output consists of dependency relation, governor term and dependent term. Dependency relation is the grammatical relation between dependent term and governor term. Governor term is the word in the sentence that the dependency relation is reported for and dependent term is the word that is dependent of the governor term. For instance, in the sentence "*Based on this, he required no operative intervention for his pseudocyst.*", *det*(*intervention-9, no-7*) *"det"* is the dependency relation, *"intervention"* is the governor term and *"no"* is the dependent term. The numbers after tokens in the parenthesis are indices of tokens with regard to their position in the sentence.

For every sentence with a concept that is considered negated by NegEx, a production chain is generated that is composed of three levels of tokens. First level token is governor of negation term, "*evidence*" in *det* (*evidence-2, No-1*). Second level tokens are dependents of first level tokens, "*of*" in *prep* (*evidence-2, of-3*). Third level tokens are dependents of second level tokens, "*dilatation*" *pobj* (*of-3, dilatation-6*). Production chain is the concatenation of these three levels of tokens, "*evidence of dilatation*". If the concept is found in the production chain, it is negated otherwise it is affirmed. The concept "*pancreatic duct dilatation"* in the sentence "*No evidence of pancreatic duct dilatation or common bile duct stones.*" is in the production chain, therefore it is negated. For concepts that are noun phrase such as "*pancreatic duct dilatation*", even if part of the noun phrase is in the production chain (dilatation), the concept is negated.

This basic rule fails in sentences with certain structures and therefore negated concepts are falsely identified as affirmed (i.e., false negative). We developed a set of rules to address the false negative results of applying DEEPEN on the IU training set. DEEPEN was developed with the mindset of decreasing the number of false positives, nonetheless we attempted to decrease the number of false negatives by addressing most common sentence structures seen in our IU training data set. Fig. 1, shows the flowchart of the algorithm used in development of DEEPEN.

Table 2 shows some examples of various rules developed in DEEPEN. More details and examples of DEEPEN rules are provided in the Appendix I. DEEPEN is written in java and is freely available for researchers to use[1].

Conjunction and rule: If there is a conjunction "*and*" in a sentence, it will be divided into two sub-sentences and negation is examined for both sub-sentences.

Preposition within rule: DEEPEN uses the collapsed representation of SDP where dependencies that involve propositions or conjunction are merged to create a direct dependency between content words. For instance, the dependencies involving prep (size-5, without-6) and pobj (without-6 inflammation-8) are collapsed into one single relation prep-without (size-5, inflammation-8). As we mentioned earlier first level token is the governor of negation term. In sentences where the negation term "*without*" is merged into the dependency relation, the governor of the relation "*prep-without*" is considered as first level token.

Preposition with/in/within rule: For propositions "*in*", "*within*", and "*with*" the SDP is only run when the concepts in these relations

[1] http://svn.code.sf.net/p/ohnlp/code/trunk/DEEPEN.

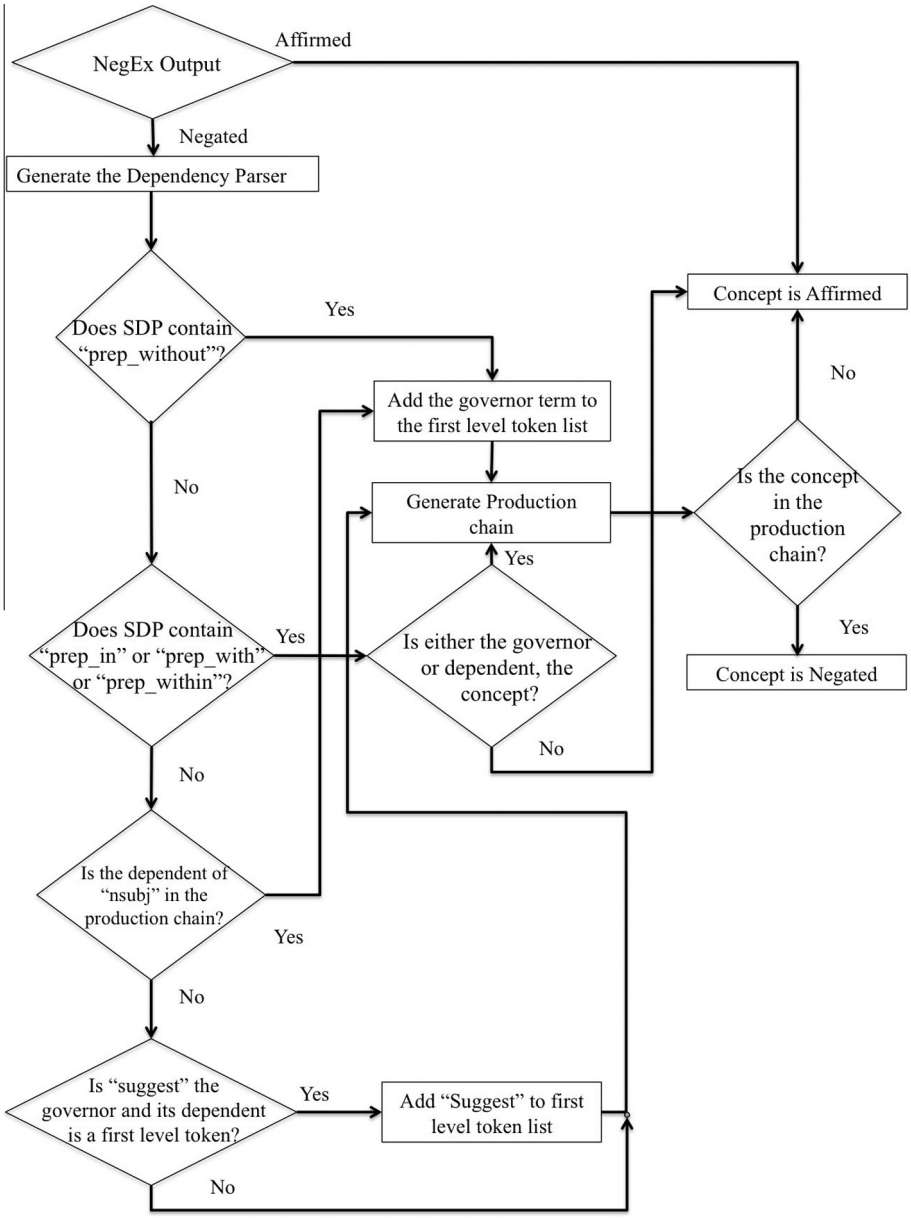

Fig. 1. Detailed flowchart of the DEEPEN algorithm.

are part of the dependent or governor terms otherwise the concept is considered as "affirmed".

Nominal subject rule: Nominal subject in SDP is a relationship in which the subject is a noun phrase such as "*No abnormally*". If the governor of this relationship is a first level token then its dependent is added to the production chain.

Suggest rule: in sentences that contain the term "*suggest*" if the dependent of the term "*suggest*" is a first level token then "*suggest*" will also be considered as a first level token.

These additional rules were added to the basic algorithm to decrease the number of incorrect assignment of present to concepts that were negated by NegEx. We stopped the development of the algorithm as we reached acceptable precision and recall of 0.9839 and 0.9983 respectively on the training set and tested the final algorithm on the test set. Identified concepts and their negation status stored in the database were exported as spreadsheet to be reviewed by two domain experts independently at IU. The inter annotator agreement between the two reviewers was 95.6%. Any discrepancies regarding the negation status of a concept was discussed with the third medical expert by looking at the complete patient report. At Mayo Clinic, we used a gold-standard dataset that has been already annotated by four annotators, further details on annotation task and schema on this dataset can be found elsewhere [39].

4. Evaluation

The system output was compared to the gold standard annotations to calculate the systems' precision, recall, and *F*-measure. Table 3 shows the relationship between the system output and manually annotated sentences.

Performance of the system is measured by precision, recall, and *F*-Measure as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (1)$$

Table 2
DEEPEN rules with relevant sentence examples and their SDP relations, concepts are shown in bold and negation terms in italic (see Appendix I for detailed dependency relations).

Rule	Sentence	Relevant dependency relations Dependency relation (governor token-index, dependent token-index)
Conjunction and	The main pancreatic duct does *not* appear disrupted and in continuity by a bridging **pseudocyst**	**pseudocyst** is affirmed in the sub-sentence "in continuity by a bridging **pseudocyst"** therefore SDP has not been applied
Preposition without	The pancreas is normal size *without* perpancreatic inflammation or **pancreatic ductal dilatation**	First level token: prep (size-5, *without*-6) Second level tokens: prep_without (size-5, inflammation-8) nsubj (size-5, pancreas-2) cop (size-5, is-3) amod (size-5, normal-4) Third level tokens: det (pancreas-2, The-1) conj_or (inflammation-8, **dilatation**-12)
Preposition in, with, and within	An abdominal CT showed a normal pancreas and gallbladder with *no* **dilated ducts**	First level token: det (**ducts**-5, *no*-3) Second level tokens: amod (**ducts**-5, **dilated**-4)
Nominal subject	No abnormally **dilated pancreatic duct**	First level token: det (abnormally-2, *No*-1) nsubj (**dilated**-3, abnormally-2)
Suggest	*No* associated fluid collection to suggest **pseudocyst** or abscess	First level token: det (collection-4, *No*-1) nsubj (suggest-6, collection-4) aux (suggest-6, to-5) dobj (suggest-6, **pseudocyst**-7) dobj (suggest-6, abscess-9) Second level tokens: amod (collection-4, associated-2) nn (collection-4, fluid-3) Third level tokens: conj_or (**pseudocyst**-7, abscess-9)

Table 3
Comparison of the system's result with manually annotated sentences.

		System output	
		True (negated)	False (affirmed)
Gold standard	True (negated)	True Positive (TP)	False Negative (FN)
	False (affirmed)	False Positive (FP)	True Negative (TN)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F\text{-Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Recall} + \text{Precision}} \quad (4)$$

5. Results

Table 4 shows the results of NegEx and DEEPEN applied to the IU and Mayo Clinic dataset. IU dataset contains 438 negated pancreatic cyst concepts (418 TPs + 20 FNs and 422 TPs + 16 FNs through NegEx and DEEPEN respectively) out of 1461 total concepts, which accounts for 30% of the data. Similarly 15.79% of disorders and 29.35% of sign and symptoms are negated in Mayo Clinic dataset. DEEPEN decreased the number of both false positives and false negatives when tested on IU dataset while it only decreased the number of false positive on Mayo Clinic dataset.

We also compared DEEPEN with DepNeg that uses dependency relations for negation detection. As the exact replication of the experiment reported in the DepNeg paper is not feasible, we compared DEEPEN's performance on the example sentences reported in the DepNeg paper. These sentences represent typical cases of DepNeg's capability of complicated negation detection as well as its limits. Table 5 shows the performance of three negation algorithms on the example sentences reported in the DepNeg paper.

DEEPEN and DepNeg could correctly identify all affirmed concepts, while DEEPEN had one less false negative than DepNeg. NegEx, however, had higher number of false positives than both DEEPEN and DepNeg while it had lower number of false negatives compared to DEEPEN and DepNeg. It should be noted that the major aim of DEEPEN and DepNeg is on having a high precision (i.e., reducing false positives).

6. Discussion

DEEPEN had higher precision and recall than NegEx on the IU dataset. However, when applied to the Mayo Clinic dataset, DEEPEN decreased false positives (i.e., higher precision) at the expense of increasing false negatives (i.e., lower recall), which resulted in lower *F*-measure than NegEx. This fact shows an interoperable issue on using heterogeneous data between institutions. NegEx uses a dictionary of negation terms that is not comprehensive. We added "lack of", "failed", "negative", "resolving" and "resolution" to NegEx's negation phrases dictionary based on observations in our training set to capture more negated concepts.

6.1. Error analysis

In what follows, we discuss some of the reasons contributed to the increasing number of false negatives.

Table 4
Comparison of DEEPEN and NegEx algorithm on IU and Mayo Clinic dataset.

		Method	TP	TN	FN	FP	Precision	Recall	*F*-Measure	Accuracy
IU dataset	Pancreatic cyst concepts	NegEx	418	983	20	40	0.9127	0.9543	0.9330	0.9589
		DEEPEN	422	1008	16	15	0.9657	0.9635	0.9645	0.9787
Mayo Clinic dataset	Disorders	NegEx	135	736	10	37	0.7849	0.9310	0.8517	0.9488
		DEEPEN	107	760	38	13	0.8917	0.7379	0.8075	0.9444
	Sign and symptoms	NegEx	113	276	10	20	0.8496	0.9187	0.8828	0.9284
		DEEPEN	95	287	28	9	0.9135	0.7724	0.8370	0.9116

Table 5
Comparison of DEEPEN, DepNeg, and NegEx, on sentences reported in the DepNeg Paper (the bold words in the sentence column denote concepts that were examined for negation status; the gray cells denote correct cases for each algorithm).

Sentence	Negation status			
	Gold standard	DEEPEN	DepNeg	NegEx
He felt that no specific therapy was available regarding **Moebius sequence**	Affirmed	Affirmed	Affirmed	Negated
I do not recommend drug treatment for **stone** prevention	Affirmed	Affirmed	Affirmed	Negated
If her **pain** should not have been resolved by that time, there is the possibility of repeating facet rhizotomy	Affirmed	Affirmed	Affirmed	Affirmed
However, I suspect that her **pain** is not due to an underlying neurologic disorder	Affirmed	Affirmed	Affirmed	Affirmed
She denies any ear pain, sore throat, odynophagia, hemoptysis, shortness-of-breath, dyspnea on exertion, chest discomfort, anorexia, nausea, weight-loss, mass, adenopathy or **pain**	Negated	Negated	Negated	Negated
Molecular fragile-X results reveal no apparent PMR-1 gene **abnormality**	Negated	Affirmed	Affirmed	Negated
Mrs. Jane Doe returns with no complaints worrisome for recurrent or **metastatic oropharynx cancer**	Negated	Affirmed	Affirmed	Negated
She is not having any incontinence or suggestion of **infection** at this time	Negated	Affirmed	Affirmed	Negated
She denies any **blood in the stool**	Negated	Negated	Affirmed	Negated

(1) Errors due to sentence detection:

Detecting the correct boundary of a sentence is a very important step in negation detection algorithm. Sentence detection in clinical notes is very challenging due to lack of end of sentence punctuation and random line breaks. Sentence detection can affect negation identification, for instance when ''HOSP NO'' and ''Diagnosis: Pancreatic pseudocyst'' in two lines were detected as one sentence the concept ''pancreatic pseudocyst'' is falsely considered negated because of the ''NO'' in ''HOSP NO'' that matches ''no'' in NegEx's negation terms. Also when multiple lines of text are considered as one sentence, dependency parser fails to correctly identify the relation between tokens in the sentence containing the concept and therefore the final negation detection result is compromised.

(2) Errors due to variations in the two institutions' corpora:

DEEPEN was developed focusing on a single concept within the IU dataset although it performed well on Mayo Clinic dataset by decreasing the number of false positive in comparison with NegEx it could not maintain the same performance consistency as tested on IU data. One of the major sentence structures in the Mayo Clinic false negatives were sentences with a negation word followed by multiple concepts separated with ''comma'' and ''or'' such as ''No associated shortness-of-breath, nausea, vomiting, diaphoresis, or light-headedness.''. All five concepts within this sentence are falsely considered affirmed by DEEPEN. More than 20 of the false negatives in sign and symptoms and 12 of false negatives in the disorders from Mayo dataset had the same structure.

(3) Conditions developed previously

Sentences that mention a condition that was previously developed in a patient but are not considered a current medical problem could be very complex and require deep contextual analysis. Following is example of two such sentences A and B from Mayo clinic and IU datasets respectively.

(A) *"Mr. X is doing very well from the standpoint of his* ***sarcoma*** *with no evidence of recurrent disease on physical examination."*
(B) *"No lesion seen at the prior site of the mid pancreatic body lesion, which was previously to represent a* ***pseudocyst****."*

Based on dependency relations, ''*sarcoma*'' and negation word ''*no*'' are not related in sentence A, however it can be inferred from the context that the concept is considered as a history and therefore negated. Likewise in sentence B, the concept ''*pseudocyst*'' is affirmed by DEEPEN because there is no relation between negation term ''No'' and the concept ''*pseudocyst*'', however previously seen pseudocyst does not mean that the patient currently has pseudocyst.

6.2. Limitations

As DEEPEN does not address the present (i.e., affirmed) concepts by NegEx. The number of concepts considered incorrectly present by DEEPEN are inherited from NegEx or due to incorrect dependency relations of SDP parsing. SDP has been created using the corpus of English web Treebank that consists of sentences from weblogs, newsgroups, etc. Therefore its performance would be lower on clinical texts that lack proper grammatical structure in comparison to general English in news and weblogs.

6.3. Future work

We are planning to address the false negative cases in Mayo Clinic dataset and also address the concepts that are affirmed by NegEx in the next release version of DEEPEN.

7. Conclusion

DEEPEN used a nested dependency relation to find out the relation between negation words and concepts to decrease the number of falsely negated concepts (i.e. false positives). It could effectively decrease the number of false positives in both the IU and Mayo

Clinic dataset in comparison with NegEx. DEEPEN shared the idea of using a dependency parser with DepNeg to find out the relation between negation words and concepts. Our approach is different from DepNeg in: (1) DepNeg does not use NegEx to find the negation status of concepts and (2) DepNeg uses rules to find out if concepts and negation words are on the same dependency path. However, DEEPEN is built on top of NegEx and only uses dependency relation rules for concepts that are negated by NegEx. The comparison of DEEPEN with DepNeg on example sentences reported in DepNeg paper showed the capability of DEEPEN in correctly identifying negation status of complicated cases.

Acknowledgments

This work was supported in part by the agency for healthcare research and quality R01 HS19818-01, grant from the office of the vice president for research at IUPUI and a joint funding from National Institute of Health R01GM102282, R01LM11369, R01LM11829, and R01 LM011934.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2015.02.010.

References

[1] Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009;42(5):760–72.
[2] Payne TD. Describing morphosyntax: a guide for field linguists. Cambridge, UK: Cambridge University Press; 1997.
[3] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. Evaluation of negation phrases in narrative clinical reports. AMIA Symp 2001:105–9.
[4] Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001;34(5):301–10.
[5] Roch AM, Mehrabi S, Krishnan A, Schmidt HE, Kesterson J, Beesley C, et al. Automated pancreatic cyst screening using natural language processing: a new tool in the early detection of pancreatic cancer. HPB 2014. http://dx.doi.org/10.1111/hpb.12375.
[6] Fundel K, Küffner R, Zimmer R. RelEx—Relation extraction using dependency parse trees. Bioinformatics 2007;23(3):365–71.
[7] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis, HLT/EMNLP. British Columbia, Canada: Vancouver; 2005. p. 347–54.
[8] Finkel J, Dingare S, Nguyen H, Nissim M, Manning C, Sinclair G. Exploiting context for biomedical entity recognition: from syntax to the web. Geneva, Switzerland: JNLPBA; 2004. p. 88–91.
[9] Cohen R, Elhadad M. Syntactic dependency parsers for biomedical-NLP. AMIA Annu Symp Proc 2012:121–8.
[10] Uzuner O, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011;18:552–6.
[11] Morante R, Daelemans W. Annotating modality and negation for a machine reading evaluation. CLEF 2011.
[12] Morante R, Sporleder C. A special issue of the computational linguistics journal on modality and negation. Comput Linguist 2010;38(2).
[13] Farkas R, Vincze V, Mora G, Csirik J, Szarvas G. The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. Uppsala 2010:1–12.
[14] Morante R, Blanco ESEM. Shared task: resolving the scope and focus of negation. SEM Montreal 2012;2012:265–74.
[15] Vincze V, Szarvas G, Farkas R, Móra G, Csirik J. The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. BMC Bioinform 2008;9(Suppl 11:S9).
[16] Pestian JP, Brew C, Matykiewicz P, Hovermale D, Johnson N, Cohen KB, et al. A shared task involving multi-label classification of clinical free text. ACL BioNLP, Workshop on BioNLP 2007. Biological, Translational, and Clinical Language Processing 2007:97–104.
[17] Morante R. Descriptive analysis of negation cues in biomedical texts. Valletta, Malta: LREC; 2010.
[18] Savova G, Masanz J, Ogren P, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507–13.
[19] Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the extensible markup language. J Am Med Inform Assoc 1999;6(1):76–87.
[20] Aronow DB, Feng F, Croft WB. Ad Hoc classification of radiology reports. J Am Med Inform Assoc 1999;6(5):393–411.
[21] Mutalik PG, Deshpande A, Nadkarni P. Use of general purpose negation detection to augment concept indexing of medical documents: a quantitative study using the umls. J Am Med Inform Assoc 2001;8:589–609.
[22] Gindl S, Kaiser K, Mik S. Syntactical negation detection in clinical practice guidelines. Stud Health Technol Inform 2008;136:187–92.
[23] Elkin PL, Brown SH, Bauer BA, Husser CS, Carruth W, Bergstrom LR, et al. A controlled trial of automated classification of negation from clinical notes. BMC Med Inform Decis Mak 2005;13(5).
[24] Sohn S, Wu S, Chute CG. Dependency parser-based negation detection in clinical narratives. AMIA Summits Transl Sci Proc 2012:1–8.
[25] Huang Y, Lowe HJ. A novel hybrid approach to automated negation detection in clinical radiology reports. J Am Med Inform Assoc 2007;14:304–11.
[26] Ballesteros M, Francisco V, Díaz A, Herrera J, Gervás P. Inferring the scope of negation in biomedical documents. New Delhi: CICLING; 2012.
[27] Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. J Biomed Inform 2009;42:839–51.
[28] Afzal Z, Pons E, Kang N, Sturkenboom M, Schuemie MJ, Kors JA. ContextD: an algorithm to identify contextual properties of medical terms in a Dutch clinical corpus. BMC Bioinformatics 2014;15:373.
[29] Skeppstedt M. Negation detection in swedish clinical text. NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, Los Angeles; 2010. p. 15–21.
[30] de Bruijn B, Cherry C, Kiritchenko S, Martin J, Zhu X. Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010. J Am Med Inform Assoc 2011;18:557–62.
[31] Cruz Díaz NP, Maña López MJ, Vázquez JM, Álvarez VP. A machine-learning approach to negation and speculation detection in clinical texts. J Am Soc Inf Sci Technol 2012;63(7):1398–410.
[32] Goldin M, Chapman WW. Learning to detect negation with 'Not' in medical texts. ACM-SIGIR; 2003.
[33] Grouin C, Abacha AB, Bernhard D, Cartoni B, Deléger L, Grau B, et al. CARAMBA: concept, assertion, and relation annotation using machine-learning based approaches. i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data, Boston; 2010.
[34] Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, et al. Negation's not solved: generalizability versus optimizability in clinical natural language processing. PLoS ONE 2014;9(11).
[35] Rea S, Pathak J, Savova G, Oniki TA, Westberge L, Beebe CE, et al. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn Project. J Biomed Inform 2012;45(4):763–71.
[36] Cairns BL, Nielsen RD, Masanz JJ, Martin JH, Palmer MS, Wh W, et al. The MiPACQ clinical question answering system. AMIA Annu Symp Proc 2011:171–80.
[37] Garla V, Lo Re 3rd V, Dorey-Stein Z, Kidwai F, Scotch M, Womack J, et al. The Yale cTAKES extensions for document classification: architecture and application. J Am Med Inform Assoc 2011;18(5):614–20.
[38] Choi JD, Palmer M. Getting the most out of transition-based dependency parsing. ACL: HLT'11, Portland, Oregon; 2011. p. 687–92.
[39] Ogren P, Savova G, Chute C. Constructing evaluation corpora for automated clinical named entity recognition. Language resources and evaluation conference. In: Proc LREC 2008:3143e50. Marrakesh, Morocco <http://www.lrec-conf.org/proceedings/lrec2008/>.
[40] de Marneffe MC, MacCartney B, Manning CD. Generating typed dependency parses from phrase structure parses. LREC 2006.
[41] de Marneffe MC, Manning CD. The Stanford natural language processing group; 2008 <http://nlp.stanford.edu/software/dependencies_manual.pdf>.